MVANet: A Unified Deep Learning Model for Binary and Multi-Class Knee MRI Diagnosis

Qi Luo[1], Chengwei Rao[2], Li Zhou[1], Si Wu[3], Jinjin Ma[4, 5]
[1]Shien-Ming Wu School of Intelligent Engineering, South China University of Technology, Guangzhou, China, [2]School of Future Technology, South China University of Technology, Guangzhou, China, [3]School of Computer Science & Engineering, South China University of Technology, Guangzhou, China
[4]School of Medicine, South China University of Technology, Guangzhou, China, [5]Institute of Future Health, South China University of Technology, Guangzhou International Campus, Guangzhou, China.

Disclosures: Authors have no disclosures

INTRODUCTION: Knee injuries are highly prevalent worldwide, and magnetic resonance imaging (MRI) is essential for their diagnosis due to superior soft-tissue contrast. Although recent studies demonstrate that deep learning (DL) achieves high performance in knee MRI analysis, its limitations include reliance on single views, a focus on binary classification, limited interpretability, and non-end-to-end designs[1, 2]. To overcome these issues, this study aims to develop an end-to-end multi-view DL model for knee MRI diagnosis and to validate its accuracy and interpretability across multiple datasets.

METHODS: This study used the open-source MRNet dataset for three binary classification tasks: ACL tear, meniscal tear, and abnormal knee diagnosis[3]. The dataset includes 1,250 exams, of which 1,130 exams from 1,088 patients were used for training and 120 exams from 111 patients for validation. Each exam provides sagittal T2-weighted, coronal T1-weighted, and axial PD-weighted sequences. For multi-class analysis, we employed the KneeMRI dataset of Štajduhar et al., consisting of 917 sagittal PD-weighted exams acquired on a Siemens Avanto 1.5-T scanner[4]. This dataset provides three ACL injury levels (non-injured, partially injured, and fully injured) as well as ACL region annotations. The annotations were used to impose a spatial constraint on the model, encouraging accurate localization of the ACL region during classification. Stratified random sampling was applied, with 120 exams reserved for validation and at least 30 cases per class included. We developed an end-to-end multi-view attention network (MVANet) composed of ResNet, transformer, and attention modules to jointly learn multi-view representations for knee MRI diagnosis. All models were initialized with ImageNet-pretrained weights and trained with data augmentation applied to the training set. Performance was assessed on the validation sets using area under the ROC curve (AUC), sensitivity, and specificity. The Dice coefficient and $IoU_{0.5}$ were employed to quantitatively evaluate the spatial agreement between CAM-derived regions and ACL ground-truth annotations.

RESULTS SECTION: On the MRNet dataset, the proposed MVANet achieved an AUC of 0.982, accuracy of 93.3%, sensitivity of 97.6%, and specificity of 84.8% for ACL diagnosis; an AUC of 0.839, accuracy of 69.2%, sensitivity of 92.3%, and specificity of 51.5% for meniscus tear diagnosis; and an AUC of 0.957, accuracy of 90.0%, sensitivity of 89.5%, and specificity of 92.0% for abnormal knee diagnosis. Using the same model, training with multi-view MRI input consistently outperformed training with a single view input: for ACL diagnosis, single-view AUCs ranged from 0.957 to 0.972; for meniscus tear diagnosis, from 0.795 to 0.819; and for abnormal knee diagnosis, from 0.886 to 0.939. Compared with baseline models, the proposed model also showed higher performance: for ACL diagnosis, AUC was 0.982 versus 0.972 (VGG), 0.938 (Vision Transformer), and 0.866 (MRNet); for meniscus tear diagnosis, AUC was 0.839 versus 0.824 (VGG16), 0.823 (Vision Transformer), and 0.782 (MRNet); and for abnormal diagnosis, AUC was 0.957 versus 0.925 (VGG16), 0.943 (Vision Transformer), and 0.907 (MRNet). On the KneeMRI dataset, the model achieved a mean AUC of 0.984 and overall accuracy of 98.4% for three-class ACL grading. For the non-injured class, accuracy was 95.0%, sensitivity 94.0%, and specificity 95.7% For the partially injured class, accuracy was 94.2%, sensitivity 96.0%, and specificity 92.9%. For the fully injured class, accuracy was 97.5%, sensitivity 85.0%, and specificity 100.0%. Incorporating CAM-based spatial constraints yielded a Dice of 0.475 and $IOU_{0.5}$ of 0.638, compared with Dice of 0.385 and $IOU_{0.5}$ of 0.541 without the constraint. As shown in Fig. 1 (B) and (C), slice-level attention weights corresponded to the annotated ACL region, and CAM visualizations localized to the annotated ACL region.

DISCUSSION: This study developed an end-to-end multi-view attention network to address limitations of DL approaches for knee MRI diagnosis. Most existing studies rely on single view[3, 5] and binary classification[6], limiting their ability to capture multi-view spatial features and assess disease severity. On the MRNet dataset, our results show that multi-view training with sagittal, coronal, and axial sequences consistently outperformed single-view training across ACL, meniscus, and abnormal knee tasks, and achieved higher AUC, sensitivity, and specificity compared with baseline models (VGG16, Vision Transformer, MRNet). On the KneeMRI dataset, MVANet achieved high accuracy for three-class ACL grading (mean AUC 0.984, accuracy up to 98.4%). Importantly, an auxiliary loss combining CAM-based constraints and slice-attention supervision was introduced to align CAM activations with ACL-annotated slices and refine slice-level weights, thereby enhancing ROI attention. This approach resembles the use of a separate ROI detector prior to diagnosis, but an end-to-end model is more convenient for clinical deployment [5, 7]. By explicitly capturing and visualizing slice-level and view-level feature weights, MVANet improves interpretability. While promising, larger multi-institutional and prospective studies are required to validate its generalizability and support clinical translation.

SIGNIFICANCE/CLINICAL RELEVANCE: This end-to-end multi-view MRI model achieved accurate knee disease diagnosis across datasets, with slice- and view-level interpretability that enhances clinical applicability of DL model.

REFERENCES: 1. Keiley et al. (2024) Eur Radiol; 2. Yu et al. (2025) Displays; 3. Bien et al. (2018) Plos Medicine; 4. Štajduhar et al. (2017) Comput Methods Programs Biomed; 5. Wang et al. (2024) Quant Img Med Surg; 6. Belton et al. (2021) MIUA; 7. Astuto et al. (2021) Radiology: AI.

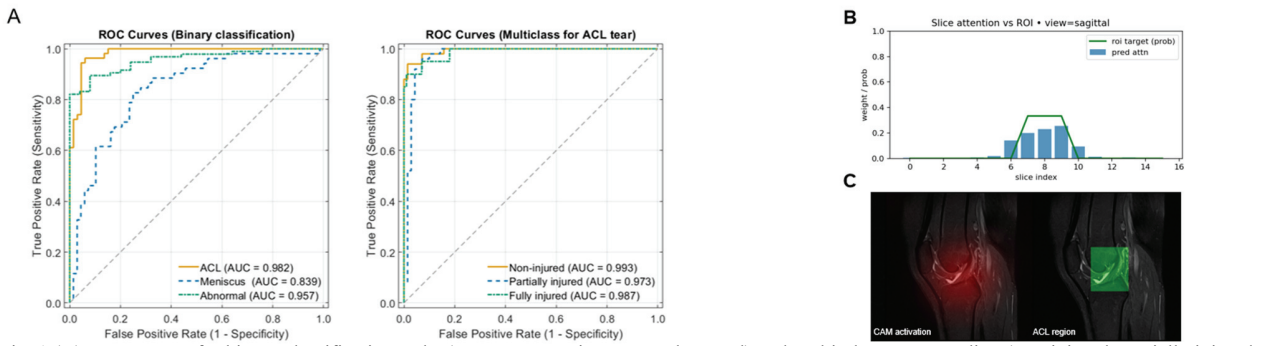

Fig. 1 (A) ROC curves for binary classification tasks (ACL tear, meniscus tear, abnormal) and multi-class ACL grading (non-injured, partially injured, fully injured). (B) Slice-level attention weights compared with ACL ROI annotations in the sagittal view. (C) Class activation map (CAM) visualization and corresponding ACL region annotation in the sagittal view.